

ABSTRACT

An input text is analyzed into morphemes by using a prescribed morphological analysis procedure to generate word strings with part-of-speech tags, including form information for parts of speech having forms, as hypotheses. The probabilities of occurrence of each hypothesis in a corpus of text are calculated by use of two or more part-of-speech n-gram models, at least one of which takes the forms of the parts of speech into consideration. Lexicalized models and class models may also be used. The models are weighted and the probabilities are combined according to the weights to obtain a single probability for each hypothesis. The hypothesis with the highest probability is selected as the solution to the morphological analysis. By combining multiple models, this method can resolve ambiguity with a higher degree of accuracy than methods that use only a single model.